they should be. It became evident that there were interactions between the various sections of the process that had not been recognized previously. This problem was investigated through studies in the smaller pilot plants. It was found that the solvent could be overhydrogenated to the point where saturates are produced faster than they are removed by cracking and distillation. High concentrations of saturates reduce solvent quality by diluting the active donor molecules. This showed that both liquefaction and solvent hydrogenation must be carefully balanced to generate the desired quantity and quality of solvent. This was another of the program's key findings and has subsequently been confirmed in ECLP. As a result, process control has been significantly improved.

Coal conversion has been 5 to 10 percent lower in ECLP than was expected from results in the smaller pilot plants. Radioactive tracer studies have shown that the residence time distributions of the reactants change significantly as the reactor diameter is increased to ECLP's 60 cm. Modeling studies are under way to correlate the data and provide a basis for scale-up to commercial size.

Status and Outlook

Our efforts in the last 15 years have demonstrated the advantages of the donor solvent approach to coal liquefaction. The EDS liquefaction process provides feed flexibility, product flexibility, high product yields, and process operability. In addition, it appears that the operability of the commercial equipment will be good, so that there will be reasonable mechanical reliability. Currently we are working on resolving the remaining issues identified in ECLP, developing a process for the liquefaction residue, and maximizing heat utilization and recovery in the preferred plant configuration. Efforts by others to complete the development of coal partial oxidation should provide technology for efficient hydrogen generation.

Note added in proof: A significant technical milestone was reached while this article was in press. It was announced at a Sponsors Management Committee meeting on 27 January 1982 that the EDS coal liquefaction technology has been successfully demonstrated and that the tests have convinced project management that the liquefaction section of a 20,000 to 30,000 ton per day, commercial-size plant operating on Illinois No. 6 coal could be designed when commercially justified. Testing on Illinois coal in the 250 ton per day pilot plant will continue at more severe conditions, including lower solvent-to-coal ratios and higher liquefaction reactor temperatures, aimed at lowering the cost of the coal liquids produced. Demonstration of the technology on Wyoming subbituminous and Texas lignite coals and development of a special boiler utilizing liquefaction bottoms are expected to be completed in 1983.

References and Notes

1. J. D. Parent, "World nonrenewable energy resources," *Energy Topics* (Institute of Gas Technology, Chicago, 27 October 1980).
2. R. J. Fiocco, paper presented at the 86th National Meeting, American Institute of Chemical Engineers, Houston, Texas, April 1979.
3. H. H. Lowry and H. J. Rose, *U.S. Bur. Mines Inf. Circ. 7420* (1947).
4. P. W. Sherwood, *Pet. Refiner* **29**, 119 (January 1950).
5. W. R. Epperly, D. T. Wade, K. W. Plumlee, *Chem. Eng. Prog.* **77**, 73 (1981).
6. R. C. Neavel, *Fuel* **55**, 237 (1976).
7. T. Aczel, *Reviews in Analytical Chemistry* (Freund, Tel Aviv, 1971), vol. 1, No. 3, pp. 226–261.
8. P. S. Maa, K. L. Trachte, R. D. Williams, paper presented at the American Chemical Society meeting, New York, August 1981.
9. W. R. Epperly, *Chemtech* **11**, 220 (1981).
10. ———, *ibid.*, p. 664.
11. L. W. Vernon, *Fuel* **59**, 45 (1980).
12. R. E. Payne, R. P. Souther, W. J. York, paper No. 27-81 presented at the 46th Midyear Refining Meeting of the American Petroleum Institute, Chicago, May 1981.
13. C. H. Montgomery, paper presented at the 65th Annual Meeting of the American Occupational Medical Association, Detroit, Mich., April 1980.
14. H. G. Shepard, *J. Soc. Occup. Med.* **31**, 9 (1981).
15. S. J. Cohen and R. E. Payne, paper presented at the Eighth Annual Conference on Coal Liquefaction, Gasification and Conversion to Electricity, Pittsburgh, Penn., August 1981.
16. R. H. Brackett, T. J. Clunie, A. M. Goldstein, paper presented at the Sixth Annual EPRI Contractor's Conference on Coal Liquefaction, Palo Alto, Calif., May 1981.
17. R. J. Platt, paper presented at Scientific Apparatus Manufacturer's Association Meeting, Washington, D.C., March 1981.
18. The work reported here is the result of a team effort of individuals too numerous to name who are or have been part of the following organizations: the Baytown Research and Development Division, the Exxon Engineering Department, and the Petroleum Research Department of Exxon Research and Engineering Company and the Synthetic Fuels Department of Exxon Co., U.S.A.

Transcriptional Control Signals of a Eukaryotic Protein-Coding Gene

Steven L. McKnight and Robert Kingsbury

Transcription of bacterial genes can be regulated by mechanisms operative at two distinct levels. One level of regulation occurs at the initiation step of RNA synthesis whereby RNA polymerase engages a bacterial gene and begins transcription. This process is governed by the interaction of specific proteins with regulatory DNA sequences located close to the site where transcription starts. Such regulatory DNA sequences can be acted on either to repress or to activate gene expression (*1*). A second level of transcriptional control in prokaryotes involves premature termination of an elongating RNA chain. This regulatory mechanism, referred to as attenuation, is mediated by a process involving intrastrand ribonucleotide sequence complementarity (*2*).

The basic physical mechanisms operative in the regulation of bacterial gene expression are likely to be used in the regulation of eukaryotic genes. It will come as no surprise, however, to find protein-DNA interaction and nucleotide sequence complementarity used in novel ways by higher organisms. For example, the phenotypic potential of different cell lineages in developing metazoan organisms become increasingly restricted as a function of embryogenesis. This phenomenon, known as embryonic determination, probably results from a mechanism that affects the implementation of semipermanent patterns of gene expression. It is not clear what specific bacterial regulatory mechanism could account for this phenomenon.

Initial studies of eukaryotic protein-coding genes, the class of genes most

The authors are employed at the Fred Hutchinson Cancer Research Center, 1124 Columbia Street, Seattle, Washington 98104.

0036-8075/82/0723-0316$01.00/0

Summary. Transcriptional control signals of a model eukaryotic protein-coding gene have been identified by a new procedure of in vitro mutagenesis. This method allows small clusters of nucleotide residues to be substituted in a site-directed manner without causing the addition or deletion of other sequences. Transcription assays of a systematic series of these clustered point mutants have led to the identification of three distinct control signals located within the 105-nucleotide residues immediately upstream from the point where transcription begins.

responsible for cellular phenotype, have focused on DNA sequences close to the site where RNA synthesis starts. While there is good evidence indicating that the DNA sequence signals which ultimately provide for the regulation of structural gene expression can lie far removed from the site where transcription begins (*3*), signals immediately upstream from the initiation sites of eukaryotic protein-coding genes at least play a role in the mechanism of gene expression (*4*).

The DNA sequences close to the start point of transcription on eukaryotic protein-coding genes probably contain at least two signals that function to facilitate transcription initiation. One signal must specify the position where RNA synthesis is to begin, and probably also determines which strand of the DNA duplex is to be transcribed, thus establishing the polarity of RNA polymerase elongation. The second signal we expect to find is one that governs the efficiency of transcription initiation. This second signal, which may well be independent of the molecular system that dictates whether a gene is transcriptionally active or not, modulates the frequency at which RNA polymerase molecules engage a gene once it has been activated. Our expectation of heritable signals that dictate discrete transcription initiation sites derives from biochemical characterizations of primary transcription products (*5*). That eukaryotic genes also contain signals which establish transcription efficiency is evidenced by visual analyses of active genes (*6*) and by biochemical measurements of the RNA synthesis rates of specific genes (*7*).

Both of these classes of transcriptional control signals have tentatively been identified in regions surrounding the transcription initiation points of eukaryotic protein-coding genes. A signal that appears to mediate a qualitative effect on gene expression by controlling the accuracy of transcription through site-specific initiation is represented by an evolutionarily conserved sequence (*8*). This signal, an A·T-rich sequence (A, adenine; T, thymine) referred to as the "TATA homology," is found between 20 and 30 nucleotides upstream from the transcription start site on most eukaryotic protein-coding genes. Correct in vitro expression of such genes by RNA polymerase II requires the TATA homology (*9, 10*). As has been noted (*10*), the nucleotide sequence of the TATA homology resembles the sequence of bacterial genes that is specifically recognized by RNA polymerase (*11*).

The second class of control sequences that appear to be required for proper expression of eukaryotic structural genes occur upstream from the TATA homology. Such upstream control sequences appear to be required for the maintenance of transcription efficiency, but tend not to affect the site specificity of transcription initiation (*12–18*). Unlike the control sequences that designate the transcription start site and have been identified as the TATA homology, the sequences that make up the second class of signals do not clearly exhibit evolutionary conservation, nor have they been precisely defined.

In an attempt to resolve the identity and functional role of transcriptional control signals closely associated with protein-coding genes, we performed a systematic mutagenesis study on a simple eukaryotic gene. The model structural gene we have analyzed is the thymidine kinase (tk) gene of herpes simplex virus. While there are a number of aspects of the tk gene that render it suitable for study, its most important qualification is that functional expression of the gene can be demonstrated unambiguously in a variety of in vivo assays (*19, 20*).

In order to study the transcriptional control signals of the tk gene we developed a method of introducing clusters of point mutations at discrete locations. These clustered point mutants have been constructed to systematically scan across a region of DNA known to harbor the transcriptional control components of the tk gene. Individual mutants were tested for the retention of transcriptional competence by a microinjection assay in living cells. These experiments identify three spatially distinct signals upstream from the tk gene that are required for in vivo expression. One such signal lies close to the 5′ terminus of the tk gene, contains the TATA homology, and mimics the function of the promoter element of bacterial genes. The other two signals of the tk gene that are crucial for expression occur ~ 50 and ~ 100 nucleotides upstream from the gene. We believe that these two distal signals represent the component ultimately responsible for governing the efficiency of tk gene expression.

Assembly of "Linker Scanning" Mutants

In order to refine our understanding of the function of the transcriptional control signals of the tk gene we sought to construct and analyze an extensive series of point mutations. Since it was unrealistic to mutate each of the 50 to 100 nucleotides thought to compose the tk transcriptional control sequences, we developed a method of mutagenesis that introduces clustered sets of point mutations at desired locations.

This method of mutagenesis requires two opposing libraries of deletion mutants that terminate deletion with the same synthetic restriction endonuclease recognition sequence in the same general region of DNA. A conventional mutagenesis method consisting of sequential treatment of linearized DNA with exonuclease III and S1 nuclease (*21*) was used to construct sets of 5′ and 3′ deletion mutants of the tk gene (*22*). The end points of 43 different 5′ deletion mutants and 42 different 3′ deletions, terminating within a 140-nucleotide segment surrounding the 5′ terminus of the tk structural gene, were identified by DNA sequencing (*23*). The end points of all 5′ and 3′ deletion mutants terminate with a synthetic Bam HI restriction site.

By recombining "matching" 5′ and 3′ deletion mutants, clustered point mutations were introduced into the tk gene. Matching, in this context, refers to two opposing deletion mutants whose deletion termini are separated by ten nucleotides. When two such mutants are recombined at the synthetic Bam HI restriction site, the ten-nucleotide residues of the linker replace the ten nucleotides that formerly separated the two deletion termini. Recombination of matching 5′ and 3′ deletion mutants does not result in either a net increase or decrease in the number of nucleotide residues in the DNA sequence.

An example of clustered point mutations introduced into the tk gene by the recombination of two matching deletion mutants is shown in Fig. 1. In this case,

eight of the ten residues between 85 and 95 nucleotides upstream from the tk structural gene are changed. Of the eight altered nucleotides, six represent transversion events, and two represent transitions. Two of the ten nucleotides (residues −95 and −90) are not changed from the "wild-type" sequence by this manipulation (Fig. 1).

A set of 18 different mutants was prepared as described above. This set was designed to systematically mutate the transcriptional control sequences of the tk gene previously localized by deletion mutagenesis (*18*). We refer to these mutants as "linker scanning" or LS mutants of the tk gene. The predicted genotype of each LS mutant was confirmed by nucleotide sequencing (Fig. 2). Of the 18 LS mutants, 15 were prepared from matching 5′ and 3′ deletions mutants. Of the three nonmatching mutants, two were prepared from parental 5′ and 3′ deletion mutants that were separated by only nine nucleotides. These LS mutants (LS −70/−61 and LS −21/−12) contain one extra nucleotide pair. The other LS mutant prepared from nonmatching parental deletions (LS −29/−18) is missing a single nucleotide pair.

This linker scanning method of mutagenesis should be applicable to essentially any purified DNA molecule. If greater or lesser mutagenic change is desired, either larger or smaller oligonucleotide linkers could be used to replace existing sequences. The major advantage of this method relative to a linker insertion method (*24*) is that, at most, it causes minimal change in the spacing between sequences. This is an important consideration in that changes in distance between transcriptional regulatory sequences will likely affect their function. Using the linker scanning method of mutagenesis we can, therefore, more directly correlate the phenotypic effects of a particular mutation with the specific nucleotide substitutions created by the linker implacement event. A certain degree of reservation should accompany such a correlation since the oligonucleotide linker might, by acting in a quasi-independent manner, spuriously affect DNA function. It is possible, for example, that the palindromic structure or unusually high concentration of CpG (C, cytosine; G, guanosine) dinucleotides of the Bam HI linker might contribute to the phenotype of different LS mutants in an inconsistent and uninterpretable manner. While such potential artifacts can not be excluded at present, our interpretations of the data will rely solely on the transition and tranversion events introduced by each LS mutant.

Development of a Pseudo-Type tk Gene

When comparing the transcriptional properties of two different DNA molecules by oocyte microinjection (*25*), we must contend with variations in the volume of DNA solution delivered into individual oocytes and the frequency at which the DNA solution is deposited into the oocyte nucleus. We assumed that the alteration in transcription efficiency affected by some LS mutants might be too subtle to be distinguished by our existing oocyte microinjection assay (*18, 20*). In order to improve the quantitative nature of the oocyte injection assay, we constructed a "marked" tk gene that is coinjected as an internal control for transcription efficiency.

A deletion mutant of the tk structural gene missing nucleotides +21 through +31 was prepared by recombining a 3′ deletion mutant that terminates at nucleotide +16 and a 5′ deletion that terminates at nucleotide +36. In addition to removing ten nucleotides coding for tk messenger RNA (mRNA) 5′ untranslated sequences, this mutant replaces an additional ten nucleotides of tk structural sequence (nucleotides +16 to +21, and +31 to +36) with that of the Bam HI linker sequence (5′-CCGGATCCGG-3′). The purpose of constructing this intragenic, +21/+31, deletion was to establish a gene that would produce an altered transcriptional product at a normal efficiency. We expected this deletion to direct the synthesis of normal levels of tk mRNA in microinjected oocytes because its mutational lesion resides in a region of DNA that is not required for transcriptional expression of the tk gene (*18*). We also expected this synthesis to be distinguishable from that of an unaltered tk gene because of the absence of ten ribonucleotide residues from the 5′ untranslated portion of the mRNA.

The method used to identify the ten-ribonucleotide deletion in mRNA de-

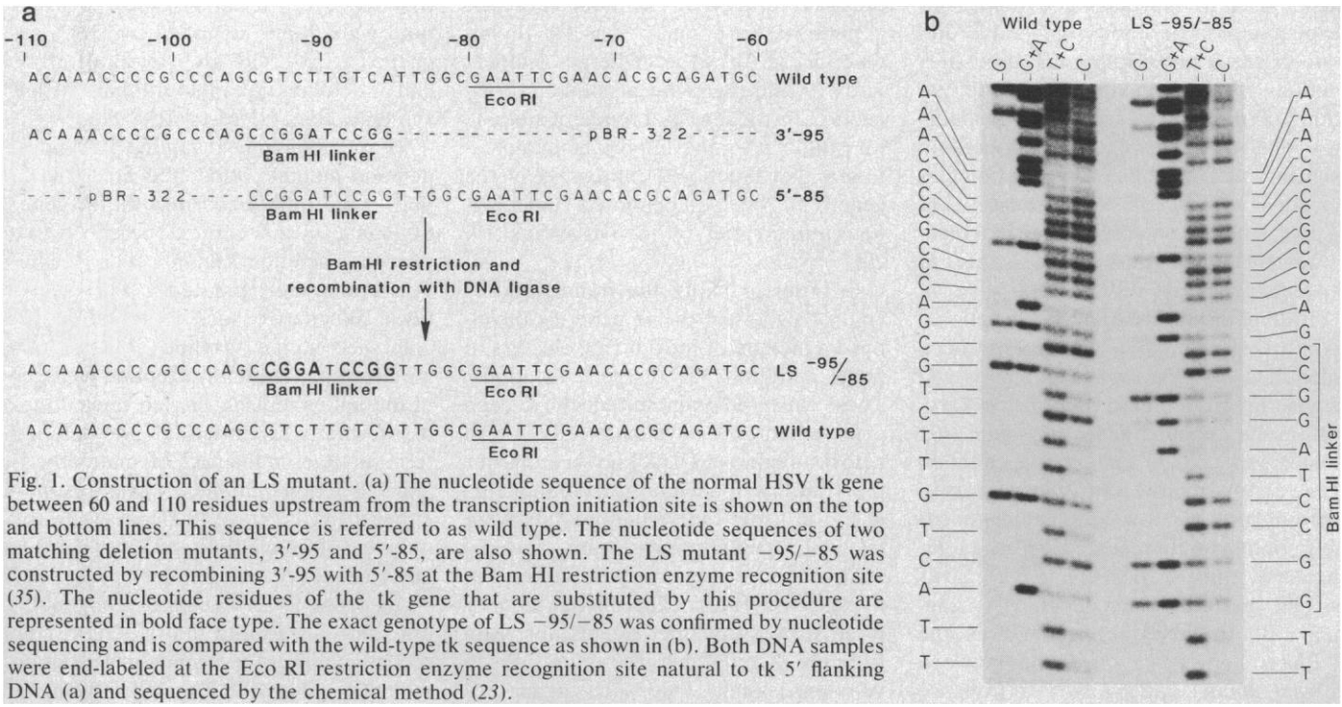

Fig. 1. Construction of an LS mutant. (a) The nucleotide sequence of the normal HSV tk gene between 60 and 110 residues upstream from the transcription initiation site is shown on the top and bottom lines. This sequence is referred to as wild type. The nucleotide sequences of two matching deletion mutants, 3′-95 and 5′-85, are also shown. The LS mutant −95/−85 was constructed by recombining 3′-95 with 5′-85 at the Bam HI restriction enzyme recognition site (*35*). The nucleotide residues of the tk gene that are substituted by this procedure are represented in bold face type. The exact genotype of LS −95/−85 was confirmed by nucleotide sequencing and is compared with the wild-type tk sequence as shown in (b). Both DNA samples were end-labeled at the Eco RI restriction enzyme recognition site natural to tk 5′ flanking DNA (a) and sequenced by the chemical method (*23*).

rived from the +21/+31 deletion mutant is a primer extension assay (Fig. 3). A single-stranded, isotopically labeled, primer DNA fragment is hybridized to tk mRNA; avian myoblastosis virus reverse transcriptase is then used to extend the primer to the 5′ terminus of the mRNA. Extension of the 40-nucleotide Bgl II–Hha I primer on tk mRNA derived from HSV-infected mammalian cells results in the synthesis of a 90-nucleotide extension product (Fig. 4). A reverse transcriptase pause site near the 5′ terminus of the authentic tk mRNA causes roughly 30 percent of the extension product to terminate four or five residues short of the 5′ terminus of the mRNA. Primer extension on tk mRNA synthesized in oocytes from a normal tk gene results in the production of exactly the same set of 85- and 90-nucleotide extension products. However, primer extension on RNA derived from oocytes injected with the +21/+31 deletion mutant of the tk gene results in the production of two extension products that measure ten nucleotides shorter than the normal set (Fig. 4). We conclude that this is a direct result of the ten-nucleotide deletion in the +21/+31 deleted tk gene.

When the normal (wild-type) tk gene and the +21/+31 deletion mutant are independently injected into frog oocytes at equivalent concentrations, they appear to be transcribed at an equivalent efficiency as evidenced by transcript mapping (Fig. 4). Coinjection of the +21/+31 deletion mutant at 50 micrograms per milliliter with wild-type DNA at 50 μg/ml also results in equivalent amounts of transcription from the two templates (Fig. 4). Since the +21/+31 deletion mutant is transcribed at an efficiency equivalent to the normal tk gene, yet makes a physically distinguishable product, we refer to it as a "pseudo wild-type" tk gene.

In order to investigate the quantitative parameters of the oocyte transcription assay, titration experiments were carried

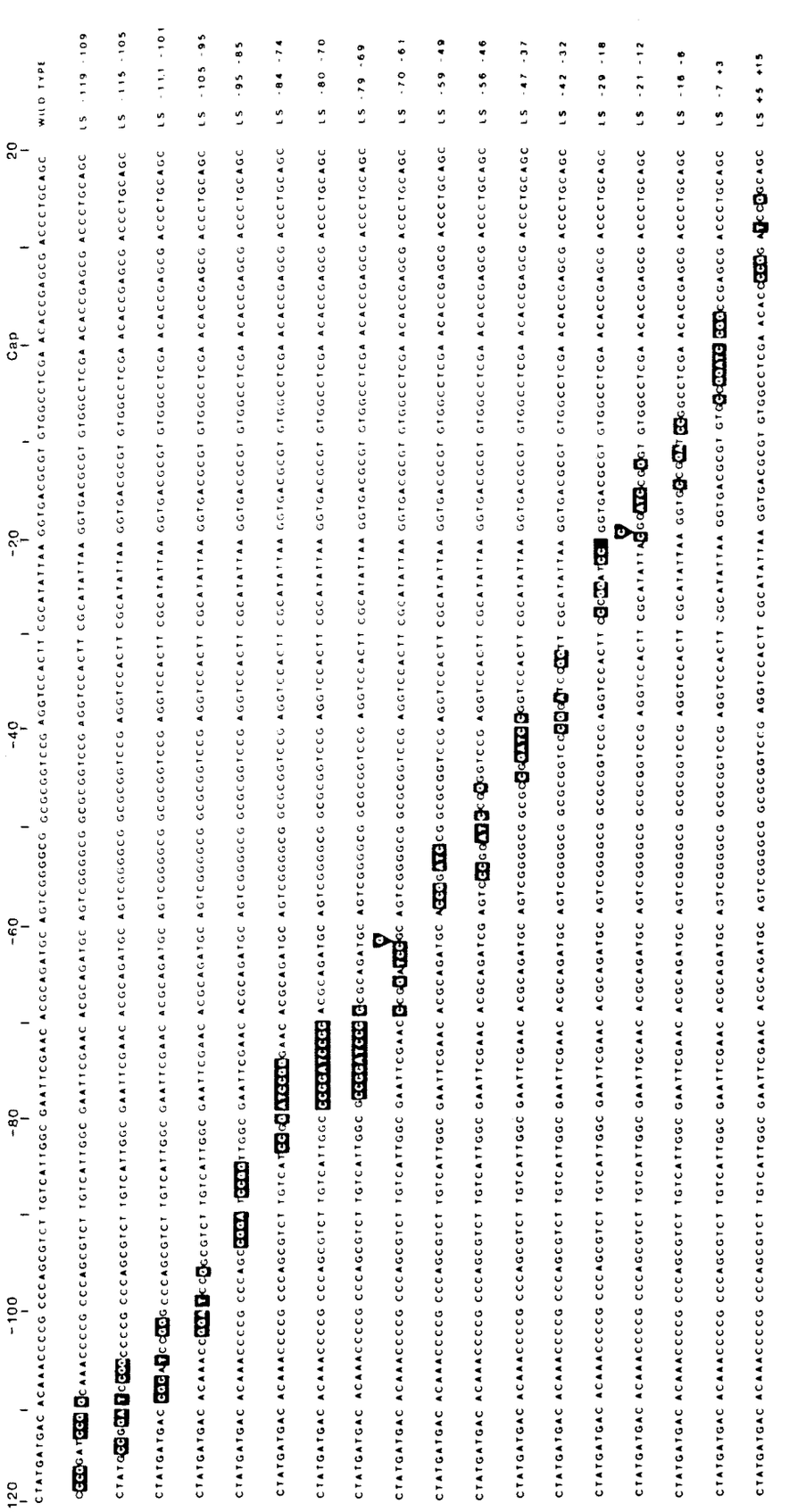

Fig. 2. Nucleotide sequences of LS mutants of the HSV tk gene. The nucleotide sequence of the wild-type thymidine kinase gene is displayed on the top line. Sequences between 120 residues upstream from the putative transcription start site (cap), and 20 nucleotides internal to the tk structural gene, are shown. The nucleotide sequences of 18 LS mutants are shown below the wild-type sequence. The nomenclature of each LS mutant is derived from the parental 5′ and 3′ deletion mutants that were recombined for its construction. Mutated nucleotide residues are displayed in the negative format. In the cases where the two parental deletions do not match (LS −70/−61, LS −29/−18, and LS −21/−12) the added or deleted residue is positioned so that the remaining residues of the Bam HI linker sequence create the minimal sum of substitutions. Each LS mutant was sequenced by the chemical method (*23*) in a region that encompassed the synthetic Bam HI linker. Nucleotide sequencing was carried out with molecules that were end-labeled at the natural Eco RI restriction enzyme recognition site at nucleotide −80. In the case of mutants that disrupt the Eco RI site (LS −84/−74, LS −80/−70, and LS −79/−69), sequencing was carried out with molecules that were end-labeled at a particular Bgl II restriction enzyme recognition site that is located 56 nucleotides internal to the tk structural gene.

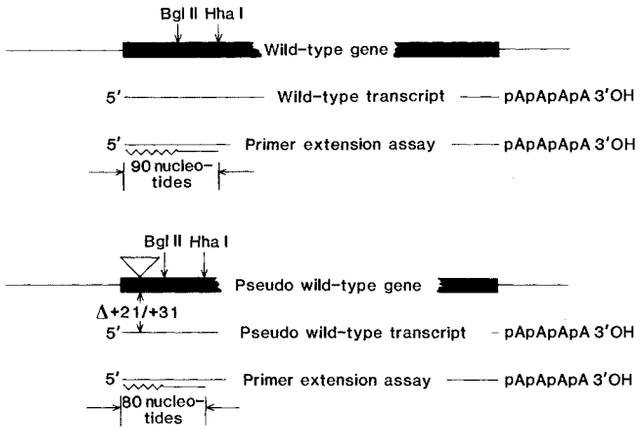

Fig. 3. Diagram of primer extension mRNA mapping assay that distinguishes between wild-type and pseudo wild-type gene products. The primer DNA molecule is prepared by double cutting tk DNA with Bgl II and Hha I restriction enzymes. The Bgl II–Hha I molecule is labeled with ^{32}P and the tk mRNA-coding strand is isolated by strand-separation gel electrophoresis (*27*). When the Bgl II–Hha I primer is hybridized to authentic tk mRNA and extended to the 5′ terminus of the mRNA with the use of avian myeloblastosis virus reverse transcriptase, a 90-nucleotide extension product is generated (upper diagram). A pseudo wild-type tk gene was prepared by deleting ten nucleotides of mRNA-coding DNA from a region of the tk gene between the Bgl II restriction enzyme recognition site and the 5′ terminus of the structural gene. When the primer extension assay is carried out on mRNA synthesized from the pseudo wild-type tk gene, an extension product ten nucleotides shorter than the normal product is generated (lower diagram).

out by injecting progressively greater amounts of psuedo wild-type tk DNA while leaving the wild-type tk gene at the original 50 μg/ml concentration. Inclusion of progressively greater concentrations of the pseudo wild-type tk gene does not result in a proportional increase in the synthesis of the pseudo wild-type transcript (Fig. 4). Although there is no more than a minimal increase in expression of the pseudo wild-type transcript as its concentration is increased, a marked reduction in the expression of the wild-type template occurs. The same result is observed in the reciprocal experiment; that is, inclusion of progressively greater amounts of wild-type template, with the pseudo wild-type gene left at a 50 μg/ml concentration, results in a reduction in pseudo wild-type transcription without a concomitant increase in wild-type transcription.

We interpret these results in the following way. Injection of 20 to 40 nanoliters of tk plasmid DNA at 50 μg/ml saturates the capacity of the oocyte to express authentic tk mRNA. No detectable increase in the synthesis of tk mRNA occurs when the number of input genes is increased. If two distinguishable, yet transcriptionally competent, tk genes are injected at an equimolar ratio, they are expressed at an equivalent efficiency. Moreover, if the ratio of the two genes is varied, the level of expression observed in the oocyte will reflect the concentration ratio of injected genes. It appears, therefore, that under our assay conditions any two microinjected tk genes compete for transcriptional expression. It is our assumption that a genotypic change in the tk gene that either increases or decreases its transcriptional competence will be reflected by its ability to compete for transcriptional expression with an unaltered tk gene in the *Xenopus* oocyte. We note, however, that since the transcriptional apparatus of the oocyte is placed under saturating conditions in this assay (Fig. 4), the phenotypic results observed for mutated tk genes may be accentuated because of internal competition by the pseudo wild-type gene.

Transcriptional Assays of Linker Scanning Mutants

Individual LS mutants were coinjected with an equimolar ratio of the pseudo wild-type tk gene into *X. laevis* oocytes. The microinjected oocytes were incubated for 24 hours, total RNA was then purified (*26*), and tk mRNA was mapped by primer extension (*27*). Since the LS mutants maintain intact structural gene sequences between the 5′ terminus of the gene and the Bgl II restriction site 56 nucleotides internal to the gene, their transcriptional products generate the full-length primer extension products (85 and 90 nucleotides). The relative transcriptional efficiency of each of the 18 LS mutants was assessed by comparing the ratio of pseudo wild-type primer extension signal to LS signal (Fig. 5). The observations presented in the form of an autoradiographic exposure (Fig. 5) were quantified by densitometric scanning (Table 1). The data obtained reveal three physically distinct regions critical to the synthesis of authentic tk mRNA in vivo.

Proximal Transcription Signals

The DNA sequences that immediately surround the putative transcription start site of the tk gene are extensively substituted by LS mutant −7/+3. This LS mutant can be expressed in amphibian oocytes in the form of a tk mRNA that is close in size to the normal gene product. Our results also suggest that transcription of LS −7/+3 can initiate at several sites upstream from the normal start point (Fig. 5). However, since the sequence of RNA made from LS −7/+3 must differ from the normal transcript, and since we have not rigorously characterized the 5′ termini of transcripts synthesized from this mutant, the role of sequences in the immediate vicinity of the transcription start site remains unresolved.

Linker scanning mutants −21/−12 and −29/−18 reveal a bona fide transcriptional control signal. This signal appears to be distinct from the start-site associated element since a LS mutant located between the two regions (LS −16/−6), which introduces five nucleotide substitutions into tk 5′ flanking DNA, exerts no detectable effect on either the accuracy or efficiency of transcription initiation. In the frog oocyte assay LS −21/−12 is capable of directing the synthesis of authentic tk mRNA, yet at a 15-fold reduced rate (Fig. 5 and Table 1). LS −28/−19 does not direct the synthesis of any detectable authentic tk mRNA when microinjected into frog oocytes. The only discernible transcriptional product derived from LS −29/−18 is a rare transcript that is roughly four nucleotides larger than the authentic mRNA (Fig. 5).

The sequences of LS −21/−12 and LS −29/−18 show alterations in the TATA homology (Fig. 2); LS −21/−12 changes the sequence from TATTAA to TATTAC, and LS −29/−18 effects multiple base substitutions in the sequence (Fig. 2). We point out that LS −21/−12 adds a single nucleotide to tk 5′ flanking DNA, and that LS −29/−18 deletes a single nucleotide. The effects of these alterations in sequence spacing cannot at present be distinguished from that of the nucleotide substitutions of these mutants. It is likely, however, that the reductions in transcription efficiency characteristic of LS −21/−12 and LS −29/−18 emphasize the involvement of the TATA homology in proper transcription initiation in vivo.

Distal Transcription Signals

Two separate DNA segments upstream from the TATA homology of the tk gene appear to play a role in maintaining quantitatively significant levels of tk mRNA synthesis in vivo. One such segment is located between 47 and 61 nucleotides upstream from the tk structural gene, the other resides between 80 and 105 nucleotides from the putative transcription start site (Fig. 5 and Table 1).

The upstream signal at −47 to −61 is distinct from the TATA control component. This is demonstrated by the fact that two mutants that lie between the two components, LS −42/−32 and LS −47/−37, show no substantial alteration in either the accuracy or efficiency of transcription despite the introduction of five- and six-nucleotide substitutions, respectively (Fig. 5 and Table 1). Moreover, the upstream signal at −47 to −61 appears to be distinct from the component at −80 to −105. This, again, is evidenced by the existence of transcriptionally competent mutants that fall between the two signals. The mutants LS −70/−61, LS −79/−69, and LS −80/−70, which substitute six, ten, and ten residues, respectively, are all accurately transcribed in microinjected frog oocytes at or near the normal efficiency (Fig. 5 and Table 1).

The two LS mutants that reveal the −47 to −61 control component are LS −59/−49 and LS −56/−46. The genotypic alterations effected by both of these LS mutants fall predominantly in a guanosine-rich segment of DNA. The two mutants are very similar in phenotype: they reduce transcriptional efficiency roughly tenfold (Table 1), yet they do not affect the accuracy of the residual transcriptional initiation events (Fig. 5).

Four LS mutations mark the −80 to −105 control component (LS −84/−74, LS −95/−85, LS −105/−95, and LS −111/−101). These four mutants are similar in phenotype to the LS mutants that mark the −47 to −61 control element in that they reduce the efficiency of transcriptional expression in vivo without affecting the position at which transcription initiates (Table 1 and Fig. 5). Mutants LS −84/−74 and LS −95/−85 reduce transcriptional efficiency to approximately 10 percent of the wild-type level, whereas mutants LS −105/−95 and LS −111/−101 effect a 20-fold reduction in transcription efficiency (Table 1). Linker scanning mutants that fall upstream from nucleotide −105 (LS −115/−105 and LS −110/−109) affect neither the efficiency nor the accuracy of in vivo expression of tk mRNA.

The genotypic alterations of LS mutants that define the −80 to −105 control component can be divided into two categories. Mutants LS −111/−101 and −105/−95, which reduce the efficiency of tk mRNA expression most dramatically, alter a cytosine (C)-rich component that reaches from nucleotide −105 to nucleotide −97 (Fig. 2). The slightly less marked reduction in transcription efficiency effected by LS −95/−85 and LS −84/−74 does not correlate with changes in this C-rich DNA segment. Instead these two mutants create multiple nucleotide substitutions in a segment of tk 5′ flanking DNA that is neither G·C nor A·T rich. The LS mutant −84/−74 alters the sequence 5′-GGCGAATT-3′, which has been identified as a conserved element of eukaryotic structural genes (*28*). It is improbable, however, that the alteration of this conserved sequence is responsible for the phenotype of LS −84/−74. This conclusion is drawn from the observation that both LS −80/−70 and LS −79/−69 alter the canonical, "CCAAT-box" sequence at a number of residues (Fig. 2) without resulting in any substantial effect on transcriptional accuracy or efficiency (Fig. 5 and Table 1).

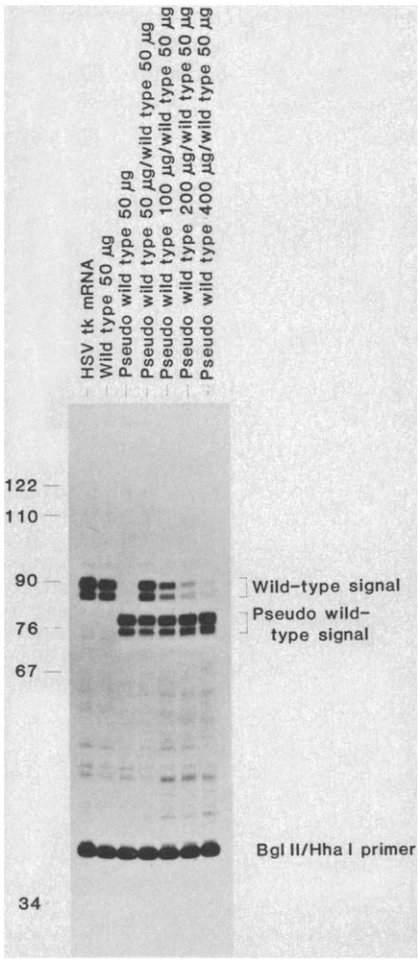

Fig. 4. Primer extension assays used to detect mRNA derived from wild-type and pseudo wild-type tk genes. An autoradiographic exposure of a polyacrylamide electrophoresis gel that was used to size labeled DNA fragments is shown. A 40-nucleotide Bgl II–Hha I primer fragment was hybridized to various tk mRNA samples then extended to the 5′ terminus of the mRNA with the use of avian myeloblastosis virus reverse transcriptase (*27*). From left to right, lane 1 shows the transcript map of authentic tk mRNA derived from HSV-infected mammalian cells. Lane 2 shows the transcript map derived from RNA isolated from *Xenopus laevis* oocytes that were microinjected with a recombinant plasmid carrying the wild-type tk gene. Individual oocytes were injected with 20 to 40 nl of a solution of covalently closed plasmid DNA (100 μg/ml). Lanes 1 and 2 exhibit a diagnostic set of 85 to 90 nucleotide extension products indicative of authentic tk mRNA. When the pseudo wild-type tk gene is injected into oocytes at a concentration of 100 μg/ml, the mRNA that was synthesized produced the transcript map shown in lane 3. The diagnostic set of extension products persists, but at a position ten nucleotides shorter than the normal set. When the pseudo wild-type tk gene is coinjected into oocyte nuclei with an equivalent concentration of the wild-type tk gene (50 μg/ml each), an equivalent amount of transcription occurs from each template (lane 4). When the concentration of pseudo wild-type gene in the injection medium is sequentially doubled, no substantial increase in its transcription occurs (lanes 5 to 7). The increased ratio of pseudo wild-type to wild-type template instead results in a proportional reduction in the expression of wild-type tk mRNA (lanes 5 to 7). Numbers to left depict positions of molecular weight marker fragments derived from digestion of pBR322 with Msp I restriction enzyme.

Implications of Mutants That Fail to Alter Transcription

Of the 18 mutants tested, nine showed no substantial phenotypic alteration. This class included one mutant that was immediately internal to the tk structural gene (LS +5/+15). Since this mutant, a deletion mutant removing nucleotides +11 to +21 (*29*), and another deletion mutant removing nucleotides +21 to +31 (the pseudo wild-type tk gene) all function normally with respect to the accuracy and efficiency of transcription initiation, it appears that the mRNA coding sequences of the tk gene do not play a significant role in regulating transcription initiation in vivo. This conclusion is consistent with interpretations drawn from deletion mapping experiments on the tk gene (*18*).

Linker scanning mutants upstream from nucleotide −105 exhibit no discernible phenotypic change. Again we recall the results of deletion mapping experiments which show that sequences more than 105 nucleotides upstream from the transcription start site are not required

for proper in vivo expression of the tk gene (*18, 29, 30*). It is not surprising, therefore, that LS −119/−109 and LS −115/−105 exhibit no effect on tk expression. Since sequences upstream from residue −105, and downstream from residue +5, are not required for appropriate tk gene expression in vivo, we conclude that all of the crucial transcriptional signals involved in the constitutive expression of tk mRNA occur within a segment of DNA no more than 110 nucleotides in length.

Linker scanning mutants that fail to affect transcription substantially were found in three separate regions within this 110-nucleotide segment of DNA: (i) between the putative site of transcription initiation and the TATA homology; (ii) between the TATA homology and a G-rich sequence located approximately 50 nucleotides upstream from the transcription start site; and (iii) between the −50, G-rich sequence and a C-rich sequence located roughly 100 nucleotides upstream from the start site.

The existence of three "mutation insensitive" segments between the most 5′ distal border of the transcription control region and the transcription start site of the tk gene is somewhat surprising. Indeed, one such region, located between 61 and 80 nucleotides upstream from the tk structural gene, harbors the conserved "CCAAT-box" homology (*28*). It may be that the regions we find between functionally critical elements are involved in subtle regulatory events that are not detected in the frog oocyte transcription assay. Alternatively, the evolutionary design of this particular structural gene may include the use of functionally inert segments between crucial transcriptional control signals. It will be of interest to assess the effects that condensation and expansion of these interstitial segments exert on the expression of tk mRNA in vivo.

Implications of Mutants That Alter Transcription

Three separate segments of DNA upstream from the 5′ terminus of the tk gene are sensitive to the introduction of clustered point mutations. The phenotypes of LS mutants that demark these "mutation-sensitive" segments can be divided into two classes. The first class of phenotypes is characterized by mutants that exhibit reduced transcription efficiency that occurs without a concomitant change in the specificity of transcription initiation. Mutants of this category have been found in three separate regions upstream from the tk structural gene: (i) a segment of DNA between 16 and 32 nucleotides upstream from the putative transcription start site; (ii) a segment between 47 and 61 nucleotides from the start site; and (iii) a segment between 80 and 105 nucleotides from the start site. A second phenotype, exhibited by LS mutant LS −29/−18, is characterized by a substantial reduction in transcription efficiency that is accompanied by an apparent alteration in the transcription initiation site.

These results support those of a deletion mutagenesis study of the tk gene (*18*); the data of that study revealed a "quantitative transcriptional control region" located between 37 and 109 nucleotides upstream from the tk structural gene and a "qualitative transcriptional control region" located between 16 and 32 nucleotides upstream from the tk gene. Our new results confirm the existence of these transcriptional control regions, and reveal several new aspects relevant to the nature of the DNA sequence signals within these control regions.

The phenotypic behavior of LS mutants −29/−18 and −21/−12 underscores the role of the sequences located between 16 and 32 nucleotides upstream from the transcription start site. This segment of tk 5′ flanking DNA, which harbors the TATA homology, appeared not to play a crucial role in transcription efficiency when assayed by deletion mapping (*18*). Although it was found that 5′ deletion mutants missing the −16 to −32 region fully lacked transcriptional function, and that one 3′ deletion mutant which terminated within the TATA homology exhibited a substantial reduction in transcriptional efficiency, two other 3′ deletions that terminated within, or up-

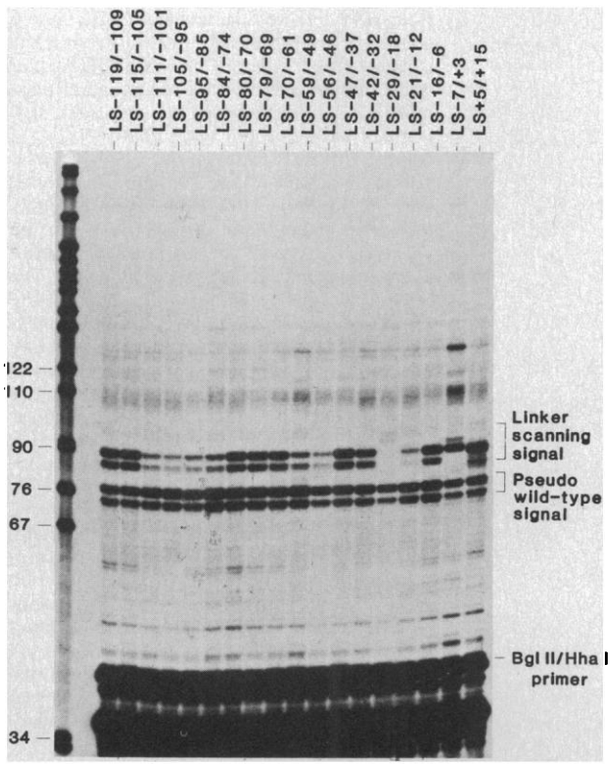

Fig. 5. Primer extension assays used to detect mRNA derived from LS mutants of the tk gene. An autoradiographic exposure of an electrophoresis gel used to size primer extension products is shown. Different samples of tk mRNA derived from microinjected oocytes were hybridized to the 40-nucleotide Bgl II–Hha I primer, then exposed to avian myeloblastosis virus reverse transcriptase (*27*). Each trial was carried out by injecting 64 oocytes with plasmid DNA at 100 μg/ml. All trials included an equimolar ratio of pseudo wild-type tk gene (50 μg/ml as injected) and LS gene (50 μg/ml as injected). Injected oocytes were incubated for 24 hours before total RNA was isolated. Primer extension assays included 10 μg of oocyte RNA (the equivalent of two oocytes) and 2×10^5 count/min of ^{32}P-labeled Bgl II-Hha I single-stranded primer DNA. Extension products migrating at 75 to 80 nucleotides reflect expression of the pseudo wild-type tk gene used as an internal control. Extension products migrating at 85 to 95 nucleotides reflect expression of the different LS mutants of the tk gene. Lanes are marked according to the identity of the LS mutant used in each injection trial. Numbers to left denote sizes in nucleotides of molecular weight marker DNA fragments.

stream from, the −16 to −32 region appeared to function at or near the normal efficiency (*18*). This discrepancy has not been resolved.

In vitro transcriptional assays clearly require an intact TATA homology for accurate transcription by RNA polymerase II (*9, 10*). In contrast, several in vivo transcriptional studies raise questions concerning the quantitative function of the TATA homology (*12, 13, 15, 18*). The results of our study indicate that both the accuracy and efficiency of in vivo transcription of the tk gene are deleteriously affected by mutations in the sequences between 16 and 32 nucleotides upstream from the putative transcription start site. We therefore conclude that the TATA homology of the tk gene and, possibly, the sequences in its immediate vicinity, represent a critical transcriptional control signal.

The TATA homology is similar in sequence (5′-TATAAA-3′) to its prokaryotic counterpart, the "Pribnow box" (5′-TATAAT-3′); and its position, compared with the Pribnow box, is located only slightly farther upstream from the transcription start site (*8, 9–11*). The *Escherichia coli* RNA polymerase directly contacts the bacterial promoter (*31*), of which the Pribnow box is an integral component. The sequence similarity and apparent functional equivalence between the −16 to −32 transcriptional signal of the tk gene and the prokaryotic promoter element lead us to believe that the site of contact between eukaryotic RNA polymerase II and the DNA of this structural gene may occur at or near to the TATA homology. We assume that such contact occurs in a manner that results in accurate transcription initiation at a discrete location.

A feature that may be peculiar to the process of transcription initiation on eukaryotic structural genes is the involvement of 5′ flanking DNA upstream from the putative RNA polymerase II contact site. In vivo assays of protein-coding eukaryotic genes have consistently demonstrated the existence of quantitatively determinative elements located at a considerable distance upstream from the TATA homology and start-site associated sequences (*12, 13, 15–18*). All of these heretofore ill-defined upstream transcriptional control signals map to regions between 50 and 400 nucleotides 5′ from the position where transcription initiation occurs. It is in this same region that the chromatin of transcriptionally active genes tends to display extreme sensitivity to various deoxyribonucleases (*32*). Although the nature of deoxyribonuclease-hypersensitive sites is not well understood, it has been proposed that these regions may be protein-depleted relative to the normal nucleosomal coverage of chromatin (*33*). If these two independently defined components, deoxyribonuclease-hypersensitive sites and quantitative control sequences, are indeed related, it may be that the primary function of quantitative control sequences is to dictate the formation of an open chromatin configuration near the initiation terminus of structural genes.

Our results delineate two critical sequence components within the upstream transcriptional control region of the tk gene. One such component is located between 47 and 61 nucleotides upstream from the putative start site, and is characterized by the guanosine-rich sequence 5′-GGGGCGGCGCGG-3′. Mutation of this sequence reduces transcription efficiency roughly tenfold. The other component of the upstream control region lies between 80 and 105 nucleotides upstream from the transcription start site. Linker scanning mutants −84/−74 and −95/−85, which mark the 3′ side of this second component, reduce transcription roughly tenfold by altering sequences that do not exhibit particularly G·C- or A·T-rich properties. On the other hand, mutants LS −105/−95 and LS −111/−101, which mark the 5′ side of this most distal transcriptional control component, reduce transcription efficiency roughly 20-fold by disrupting the C-rich sequence 5′-CCCCGCCC-3′ (Fig. 2).

There is strong evidence that the C-rich sequence located between 105 and 97 nucleotides upstream from the tk gene is critical to transcription. For example, we know from deletion mutagenesis that nucleotide sequences upstream from residue −105 are not required for efficient transcription in vivo (*18, 30*). Since the only mutational lesions of LS −111/−101 that occur downstream from residue −105 are the two C to G transversions at residues −102 and −103 (Fig. 2), we believe that they alone elicit the 20-fold reduction in transcription efficiency characteristic of this mutant.

It is interesting that mutants in the guanosine-rich segment of tk 5′ flanking DNA (LS −59/−49 and LS −56/−46) lead to a phenotype similar to that of mutants that alter the C-rich component (LS −111/−101 and LS −105/−95). We have noted that a perfectly complementing, six base pair inverted repeat occurs within these components. Were this inverted repeat to bring about an intrastrand interaction between the G-rich element at −50 and the C-rich element at −100, a 42-nucleotide loop would extend from both DNA strands in the region separating the two elements.

The four LS mutants that are most deleterious to the function of the upstream control region of the tk gene all substitute nucleotide residues within the six complementing pairs of this inverted repeat. Moreover, no LS mutant or deletion mutant of the tk gene has been found that alters this inverted repeat without concomitantly reducing transcriptional efficiency. While we speculate that an intrastrand interaction between these inverted repeats may be required for efficient in vivo tk gene expression, we stress that we have made no direct observations to support this hypothesis. We also recall that the sequences between 84 and 95 nucleotides upstream from the tk structural gene, which would not complement other tk sequences in this hypothetical stem-loop structure, are clearly crucial to the function of the upstream transcriptional control region (Fig. 2 and Table 1).

We propose the following model. The upstream control region of the tk gene facilitates initial entry of RNA polymerase II onto the tk gene. This event allows contact to be made between RNA polymerase and TATA. Such contact results

Table 1. Transcription efficiencies of LS mutants. The relative transcription efficiency of each linker scanning mutant was quantified by densitometric scanning of an autoradiographic exposure shown in Fig. 5. Each lane of the exposure was scanned by means of a soft laser densitometric scanner. The dual optical density peaks derived from the pseudo wild-type transcription signal were compared for each experiment, with the peaks assumed to derive from the LS transcription signal. Expression efficiency refers to the ratio of LS transcription signal to pseudo wild-type signal, with 1.0 reflecting equivalence.

Mutant	Expression efficiency
LS −119/−109	1.20
LS −115/−105	1.36
LS −111/−101	0.04
LS −105/−95	0.03
LS −95/−85	0.12
LS −84/−74	0.09
LS −80/−70	0.87
LS −79/−69	0.70
LS −70/−61	0.94
LS −59/−49	0.08
LS −56/−46	0.11
LS −47/−37	1.27
LS −42/−32	0.68
LS −29/−18	0.06*
LS −21/−12	0.07
LS −16/−6	0.96
LS −7/+3	1.04*
LS +5/+15	1.18*

*Samples where the linker scanning transcript mapping signal was derived from a region of the electrophoresis gel that did not perfectly match the position of the wild-type transcript map.

in transcription initiation at a specific site. The efficiency of the RNA polymerase entry step is dictated by the genotype of the upstream control region. This component of a structural gene would, therefore, be assumed to constitute an inherent modulatory signal that plays an important role in governing the ultimate abundance of the gene product within the cell.

There are several schemes that can be proposed to account for the initial entry of RNA polymerase. First, the event could result from the recognition of an open chromatin complex somehow dictated by the upstream control sequences. Such a model would not strictly require that RNA polymerase recognize the upstream control element in a sequence-specific manner. The open complex might simply serve as an entry point for the RNA polymerase to begin a scan for a sequence-specific recognition site such as TATA. Alternatively, RNA polymerase entry might result from specific recognition of upstream control sequences. A variation on this second scheme would involve the function of a transcription factor that maintains specific affinity for both the upstream control sequences and the RNA polymerase molecule.

Finally, we think that the upstream control region of the tk gene might function as a consequence of a change in its conformation from the normal B-form duplex to, for example, an intrastrand stem-loop configuration. In this regard we make special note of the recent results of Larsen and Weintraub (*34*) which suggest that the deoxyribonuclease-hypersensitive sites of active structural genes contain single-stranded DNA. Resolution of these fundamental questions with the use of a simple, constitutively expressed eukaryotic structural gene may lead to a strategy for studying other questions of temporal and tissue-specific gene regulation in higher organisms.

References and Notes

1. J. H. Miller and W. S. Reznikoff, Eds., *The Operon* (Cold Spring Harbor Laboratory, New York, 1978).
2. C. Yanofsky, *Nature (London)* **289**, 751 (1981).
3. A. J. S. Klar, J. N. Strathern, J. R. Broach, J. B. Hicks, *ibid.*, p. 239; K. A. Nasmyth, K. Tatchell, B. D. Hall, C. Astell, M. Smith, *ibid.*, p. 244; T. Maniatis, E. F. Fritsch, J. Lauer, R. M. Lawn, *Annu. Rev. Genet.* **14**, 145 (1980).
4. R. Breathnach and P. Chambon, *Annu. Rev. Biochem.* **50**, 349 (1981).
5. E. B. Ziff and R. M. Evans, *Cell* **15**, 1463 (1978); M. Tsuda, Y. Ohshima, Y. Suzuki, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 4872 (1979); R. Contreras and W. Fiers, *Nucleic Acids Res.* **9**, 215 (1981).
6. S. L. McKnight and O. L. Miller, *Cell* **8**, 305 (1976); V. E. Foe, L. E. Wilkinson, C. D. Laird, *ibid.* **9**, 131 (1976); C. D. Laird, L. E. Wilkinson, V. E. Foe, W. Y. Chooi; *Chromosoma (Berlin)* **58**, 169 (1976); S. L. McKnight and O. L. Miller, *Cell* **17**, 551 (1979); S. Busby and A. H. Bakken, *Chromosoma (Berlin)* **71**, 249 (1979); *ibid.* **79**, 85 (1980).
7. M. H. Harpold, R. M. Evans, M. Salditt-Georgieff, J. E. Darnell, *Cell* **17**, 1025 (1979); E. Derman *et al.*, *ibid.* **23**, 731 (1981).
8. M. Goldberg, thesis, Stanford University (1979).
9. J. Corden, B. Wasylyk, A. Buchwalder, P. Sassone-Corsi, D. Kedinger, P. Chambon, *Science* **209**, 1406 (1980).
10. B. Wasylyk *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 7024 (1981); G. C. Grosveld, C. K. Shewmaker, P. Jat, R. A. Flavell, *Cell* **25**, 215 (1981); S. L. Hu and J. L. Manley, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 820 (1981); Y. Tsujimoto, S. Hirose, M. Tsuda, Y. Suzuki, *ibid.*, p. 4838.
11. D. Pribnow, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 784 (1975); H. Schaller, C. Gray, K. Herrmann, *ibid.*, p. 737.
12. R. Grosschedl and M. L. Birnstiel, *ibid.* **77**, 1432 (1980).
13. ———, *ibid.*, p. 7102.
14. R. M. Meyers, D. C. Rio, A. K. Robbins, R. Tjian, *Cell* **25**, 373 (1981).
15. C. Benoist and P. Chambon, *Nature (London)* **290**, 304 (1981).
16. P. Dierks, A. Van Oyen, N. Mantei, C. Weissman, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 1411 (1981).
17. P. Mellon, V. Parker, Y. Gluzman, T. Maniatis, *Cell* **27**, 279 (1981).
18. S. L. McKnight, E. R. Gavis, R. Kingsbury, R. Axel, *ibid.* **25**, 385 (1981).
19. M. Wigler, S. Silverstein, L. Lee, A. Pellicer, Y. Chang, R. Axel, *ibid.* **11**, 223 (1977); N. J. Maitland and J. K. McDougall, *ibid.*, p. 233; M. R. Capecchi, *ibid.* **22**, 479 (1980).
20. S. L. McKnight and E. R. Gavis, *Nucleic Acids Res.* **8**, 5931 (1980).
21. S. Sakonju, D. F. Bogenhagen, D. D. Brown, *Cell* **19**, 13 (1980).
22. Recombinant DNA derived from the macroplaque strain of herpes simplex virus type 1 was enzymatically digested in the region surrounding the 5′ terminus of the tk gene (thymidine kinase gene) by the combined use of exonuclease III and S1 nuclease. All deletion termini were blunt-end ligated to the synthetic oligonucleotide 5′-CCGGATCCGG-3′, which contains the recognition sequence for Bam HI restriction enzyme. Approximately 100 5′ deletion mutants and 75 3′ deletion mutants were diagnosed by gel electrophoresis to maintain a deletion terminus between nucleotide residue −120 and residue +20 of the tk gene. Nucleotide residue +1 refers to the putative site of transcription initiation. Each mutant was sequenced by chemical methods to identify the deletion end point precisely.
23. A. Maxam and W. Gilbert, *Methods Enzymol.* **65**, 499 (1980).
24. F. Heffron, M. So, B. J. McCarthy, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 6012 (1978).
25. J. B. Gurdon, *The Control of Gene Expression in Animal Development* (Harvard Univ. Press, Cambridge, Mass., 1974); J. E. Mertz and J. B. Gurdon, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 1502 (1977); A. Kressman, S. G. Clarkson, J. Telford, M. L. Birnstiel, *Cold Spring Harbor Symp. Quant. Biol.* **42**, 1077 (1977); J. B. Gurdon and D. D. Brown, *Dev. Biol.* **67**, 346 (1978).
26. E. Probst, A. Kressman, M. L. Birnstiel, *J. Mol. Biol.* **135**, 709 (1979).
27. RNA prepared from microinjected oocytes was assayed for the presence of authentic tk mRNA and mRNA derived from transcription of the pseudo wild-type tk gene by a primer extension method with the use of avian myeloblastosis virus (AMV) reverse transcriptase. An isotopically labeled DNA probe was prepared by digesting recombinant tk DNA with Bgl II endonuclease, resecting the 3′ strand from the Bgl II site with exonuclease III, and filling in the resected molecules with [α-^{32}P]dCTP (deoxycytidine triphosphate) and [α-^{32}P]dTTP (deoxythymidine triphosphate) (300 to 500 Ci/mmole) and AMV reverse transcriptase. The labeled DNA molecules were then digested with Hha I endonuclease to liberate a 40-nucleotide duplex fragment, and the strands were separated by gel electrophoresis. The tk coding strand of this fragment was recovered and used as a hybridization primer on tk mRNA synthesized in microinjected frog oocytes. Hybridization reaction (20 μl) consisting of oocyte RNA (10 to 20 μg) and the Bgl II–Hha I primer DNA fragment labeled with ^{32}P (2×10^4 to 4×10^4 count/min) were incubated for 1 to 2 hours at 60°C in 0.25*M* KCl, 10 m*M* tris-HCl, *p*H 7.5, and 1 m*M* EDTA. Reactions were cooled to room temperature and diluted with a reverse transcriptase reaction mixture with final concentrations of 75 m*M* KCl, 0.25 m*M* EDTA, 10 m*M* $MgCl_2$, 20 m*M* tris-HCl (*p*H 8.3), 10 m*M* dithiothreitol, 0.25 m*M* deoxyribonucleoside triphosphates, actinomycin D (100 μg/ml), and AMV reverse transcriptase (100 unit/ml). Primer extension reactions were incubated at 37°C for 1 hour, precipitated with ethanol, washed, and resuspended in 20 μl of 99 percent formamide containing 0.1 percent bromphenol blue and xylene cyanol. Samples were then subjected to electrophoresis through 0.4-mm sequencing gels consisting of 8 to 12 percent polyacrylamide polymerized in 8*M* urea. Primer extension products were visualized on autoradiographic exposures of electrophoresis gels. The radioactivity in DNA bands was ascertained by means of densitometric scanning with a soft laser.
28. C. Benoist, K. O'Hare, R. Bretnach, P. Chambon, *Nucleic Acids Res.* **8**, 127 (1980); A. Efstratiadis *et al.*, *Cell* **21**, 653 (1980).
29. S. L. McKnight and R. C. Kingsbury, unpublished results.
30. A. El Kareh, M. Ostrander, S. Silverstein, J. Smiley, in preparation.
31. L. Johnsrud, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 5314 (1978).
32. C. Wu, *Nature (London)* **286**, 854 (1980); J. Stalder, A. Larsen, J. D. Engel, M. Dolan, M. Groudine, H. Weintraub, *Cell* **20**, 451 (1980); C. Wu and W. Gilbert, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 1577 (1981); M. Groudine, R. Eisenman, H. Weintraub, *Nature (London)* **292**, 311 (1981); R. Shermoen and S. Beckendorf, *Cell*, in press.
33. A. J. Varshavsky, O. H. Sundin, M. J. Bohn, *Cell* **16**, 453 (1979); J. D. McGhee, W. I. Wood, M. Dolan, J. D. Engel, G. Felsenfeld, *ibid.* **27**, 45 (1981).
34. A. Larsen and H. Weintraub, *ibid.*, in press.
35. Matching 5′ and 3′ deletion mutants, as described in Fig. 1, were recombined at the site of the synthetic Bam HI linker as follows. Purified 5′ deletion mutant plasmid DNA (5 μg) was mixed with purified plasmid of a matching 3′ deletion mutant (5 μg). The mixture was subjected to restriction digestion with Bam HI endonuclease which digests each of the two plasmids to the linear form at the site of the synthetic Bam HI linker. Linearized DNA was precipitated with ethanol, resuspended in 25 μl of a buffer containing 50 m*M* tris-HCl, *p*H 7.5, 10 m*M* Mg_2Cl, 1 m*M* ATP, 5 m*M* dithiothreitol, and ligated overnight at 4°C with T4 DNA ligase (10 unit/ml). Ligated samples were heated to 60°C for 15 minutes to inactivate the ligase, then digested with Hind III. The DNA was then subjected to electrophoresis on a 1 percent agarose gel, and a 2.1 kilobase Hind III generated DNA fragment was recovered. This fragment, which represents the recombination product of matching 5′ and 3′ deletion mutants, was inserted into the Hind III site of pBR322 by ligation, and the resultant recombinant plasmid was cloned in *Escherichia coli* strain HB-101. Each LS mutant thus constructed was inserted into the plasmid in the same orientation. Purified plasmid DNA of each mutant was chemically sequenced to confirm the predicted genotype.

36. Supported by the Helen Hay Whitney Foundation for Medical Research, The Carnegie Institution of Washington, and the National Institute of Arthritis and Infectious Diseases. We thank D. Brown, D. Bogenhagen, M. Wormington, M. Dunaway, R. Harland, B. Torok-Storb, S. Eisenberg, H. Weintraub, M. Groudine, and R. Axel for contributions made during the various stages of this project.